

A DATA LAKE IS NOT A CDR AND A CDR IS NOT A DATA LAKE...

*UNDERSTANDING THE DIFFERENCES AND THE
VALUE OF EACH*

Chris Decker, DH05
EU Connect 2023



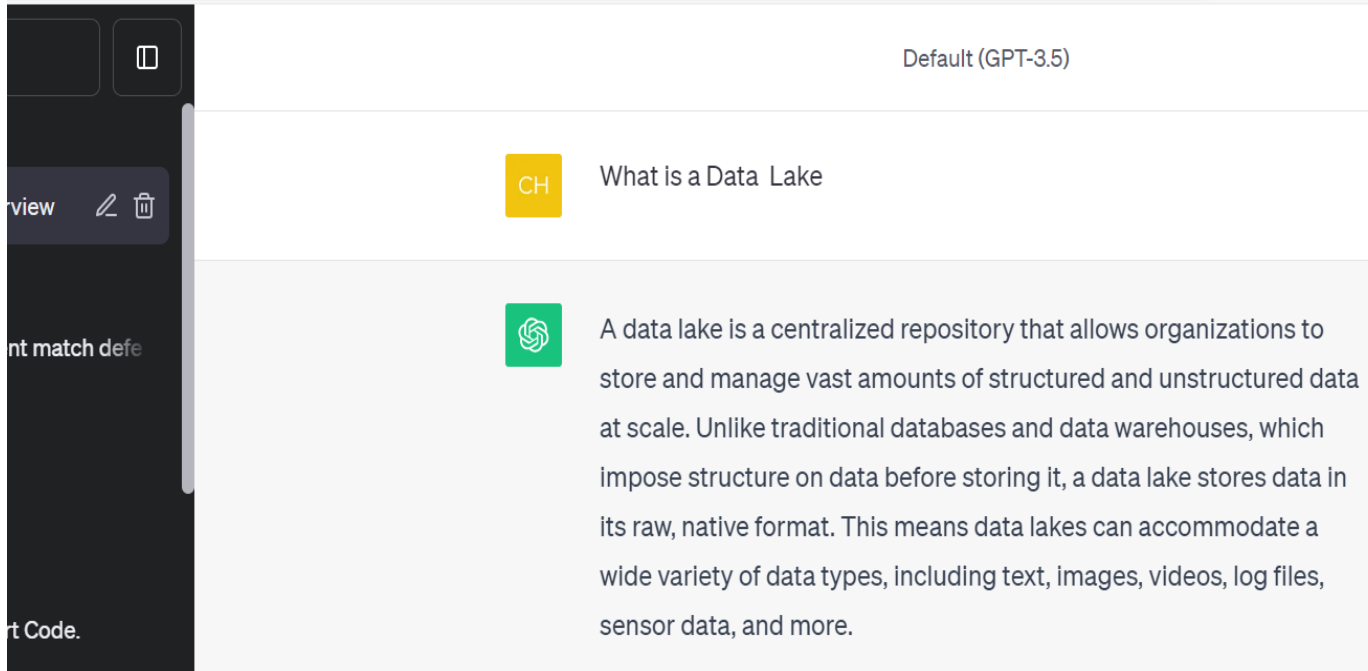
AGENDA

- What is a Data Lake?
- What is a CDR?
- Data Lake vs CDR: Similarities and Differences
- Leveraging both to Accelerate Clinical Analytics
- What does the Future Hold?



WHAT IS A DATA LAKE?

<https://chat.openai.com/c/88d6cb68-cf39-4440-8db0-4e16f770fe15>



A DATA LAKE IS...

- ✓ Repository storing large amounts of structured and unstructured data
- ✓ Information stored in native format accommodating a large variety of data types
- ✓ Flexible, scalable, and cost-effective storage



KEY CHALLENGES....

- Managing the data quality across such a diverse set of data
- Variability in the definition and management of metadata
- Complex data governance and protecting your data

WHAT IS A CDR (CLINICAL DATA REPOSITORY)?

<https://chat.openai.com/c/88d6cb68-cf39-4440-8db0-4e16f770fe15>

Default (GPT-3.5)



What is a Clinical Data Repository



A Clinical Data Repository (CDR) is a centralized database or digital storage system that is designed to collect, store, and manage clinical and healthcare-related data from various sources. CDRs play a crucial role in the healthcare industry by providing a structured and secure platform for storing patient health records, medical data, and other relevant information. Here are some key features and purposes of a Clinical Data Repository:

1. **Data Integration:** CDRs aggregate data from a wide range of sources within a healthcare organization, including electronic health records (EHRs), laboratory results, radiology reports, billing information, and more. This integration allows for a comprehensive view of patient health.
2. **Data Standardization:** Healthcare data comes in various formats and standards. CDRs often standardize data by converting it into a common format or standard terminology, making it easier to analyze and exchange data across different systems.

A CDR IS A...

- ✓ File based repository to store and manage clinical data and artifacts
- ✓ Heavily controlled and secure environment to support the right access to the right user
- ✓ Focused on standardization and enabling the analytics workflow required for submission

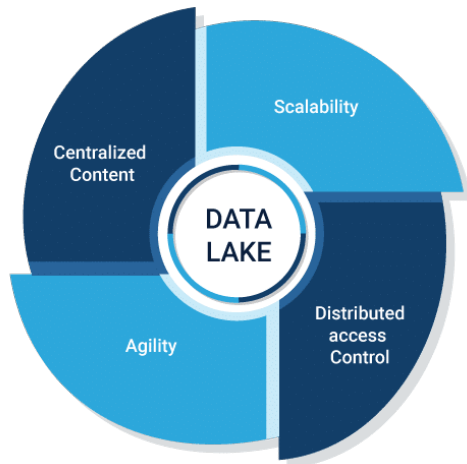


KEY CHALLENGES....

- Increase in volume/diversity of data not always supported in traditional CDR
- Controls and traceability get in the way of the workflow
- Legacy artifacts and processes keep us stuck – transforming data 5 times and creating A4 (or 8.5x11) pieces of paper (i.e. TLFS, TFLs, TLGs, TGLs...)

DATA LAKE VS. CDR

CAPABILITY	DATA LAKE	CDR
Data Types	Unstructured/Structured and anywhere in between	Structured & File Based
Data Schema	Raw, unfiltered, and easily updated	Processed, Derived, and difficult to evolve
Users	Data scientists, data engineers	Programmers, Statisticians
Security	Distributed, data point driven, easily complex	File based, groups/users
Workloads	Analytical and Exploratory	Operational and Transactional

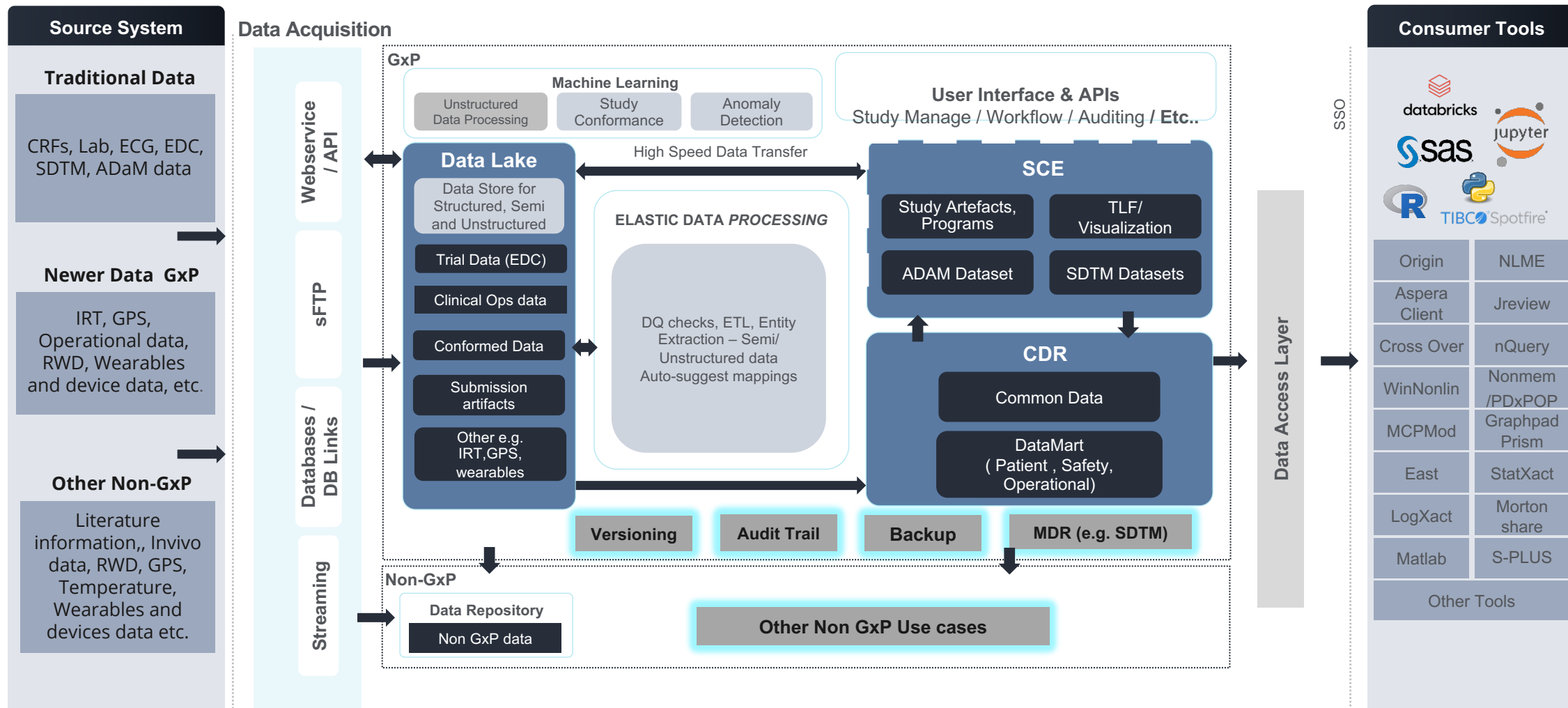


COMMON QUESTIONS....

- *Where do you standardize?*
- *Where do you transform the data?*
- *Does all data go through the Data Lake?*
- *When and what data produced in the CDR goes back in Data Lake?*



CASE STUDY: LEVERAGING BOTH TO ACCELERATE YOUR WORK





What Does the Future Hold?

TODAY'S LIMITATIONS



Why do we use
a CDR?

That's an easy question: We stick data in
rows and columns so we need folders, files,
and lots of transformations



"In the real world, we want to capture the complexity and diversity of the standard of care without forcing it into an artificial structure. The current technology framework used within clinical trials to collect information often makes researchers constrain data to a structure that is not real and has no context..... and forces them into a rigid non intuitive thought process"

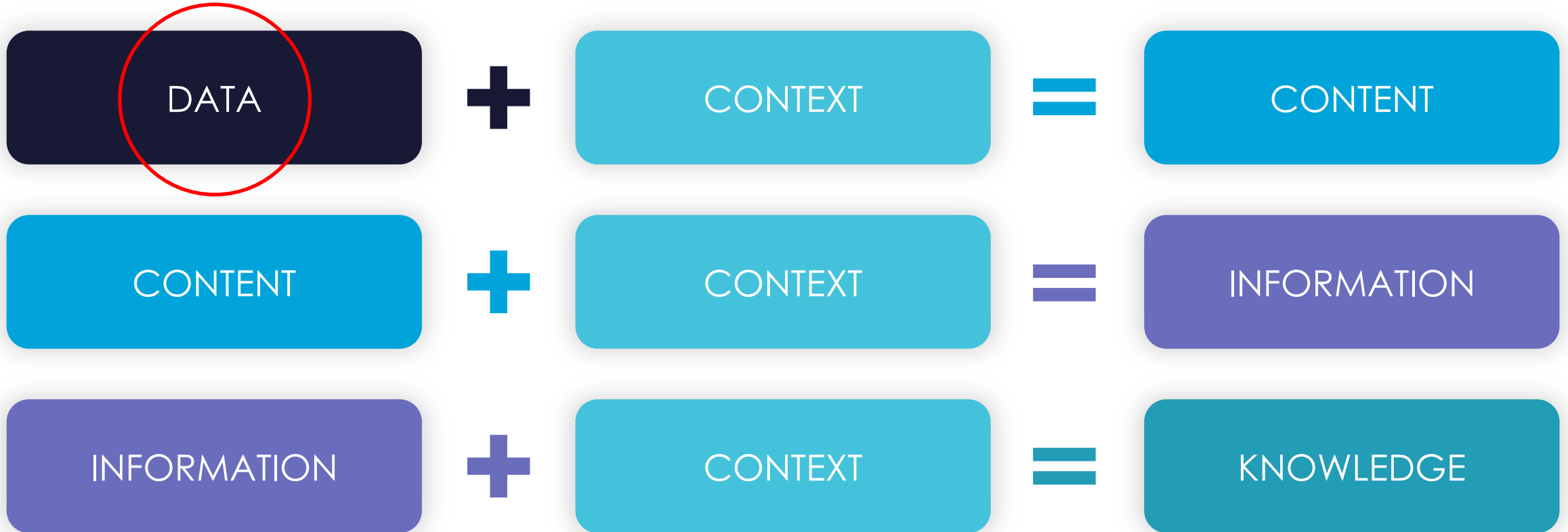
FLORENCE BARKATS, RWE SME



*“We are drowning in
information but starved
for knowledge.”*

JOHN NAISBITT

CHANGING THE DATA PARADIGM



#KillRowsAndColumns | #killDeathByTransformation | #killTLFs



*"The reasonable man adapts himself to
the world; the unreasonable one
persists in trying to adapt the world to
himself. Therefore, all progress depends
on the unreasonable man."*

GEORGE BERNARD SHAW

IT'S TIME TO BE UNREASONABLE...



STATEMENT OF PURPOSE

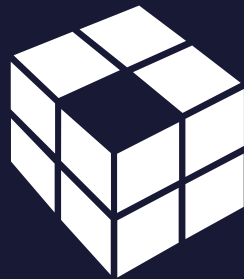
To help our clients bring life enhancing
products to market faster

Information Solutions for Life

QUESTIONS?

CHRIS DECKER

chris.decker@instem.com



Instem

instem.com | info@instem.com